DOCUMENT RESUME

ED 411 836 IR 056 663

AUTHOR Nuyts, Annemie; Jonckheere, Charles
TITLE Automated Patent Searching in the EPO: From Online Searching to Document Delivery.
PUB DATE 1996-00-00
NOTE 8p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; Automation; Computer Software; Databases; *Document Delivery; Foreign Countries; Full Text Databases; *Information Services; Information Storage; Information Technology; *Online Searching; Online Systems; *Patents; Storage; Technological Advancement; *Technology Transfer
IDENTIFIERS Browsing; *European Patent Office; Interactive Communication; Patent Office

ABSTRACT

The European Patent Office (EPO) has recently implemented the last part of its ambitious automation project aimed at creating an automated search environment for approximately 1200 EPO patent search examiners. The examiners now have at their disposal an integrated set of tools offering a full range of functionalities from online searching, via full text browsing, to document delivery on the workstation screen or a nearby departmental printer. Online searching, via a common command language, is carried out in a great number of bibliographic and full-text databases on the EPO in-house host service or in commercial databases. Specially developed "dual-mode" viewer software enables the examiner easily to browse full-text documents (in character mode) and view patent images (in facsimile mode). Browsing and viewing, as online searching, are interactive operations with very short response times. The recently installed electronic document server, finally, gives access to the EPO's complete search collection comprising some 25 MIO documents, totaling approximately 14 terabytes of storage space. The storage technology is based on the use of magnetic cartridges placed in robots, which offer not an interactive online service, but rather a "near-line" service with maximum 15 minutes response time. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 836

Automated Patent Searching in the EPO: From Online Searching to Document Delivery

By:

Annemie Nuyts

Charles Jonckheere

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B.P. Jeapes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR056663

BEST COPY AVAILABLE

Automated patent searching in the EPO: from online searching to document delivery

Annemie Nuyts and Charles Jonckheere
European Patent Office, The Netherlands

Abstract: *The European Patent Office (EPO) has recently implemented the last part of its ambitious automation project aimed at creating an automated search environment for approximately 1200 EPO patent search examiners. The examiners now have at their disposal an integrated set of tools offering a full range of functionalities from online searching, via full text browsing, to document delivery on the workstation screen or a nearby departmental printer.*
Online searching, via a common command language, is carried out in a great number of bibliographic and full-text databases on the EPO in-house host service or in commercial databases (on STN, DIALOG, QUESTEL etc.). Specially developed 'dual-mode' viewer software enables the examiner easily to browse full-text documents (in character mode) and view patent images (in facsimile mode). Browsing and viewing, as online searching, are interactive operations with very short response times. The recently installed electronic document server, finally, gives access to the EPO's complete search collection comprising some 25 MIO documents (250 MIO pages compressed according to CCITT group 4.T6), totalling approximately 14 terabytes of storage space. The storage technology is based on the use of magnetic cartridges placed in robots, which offer not an interactive online service but rather a 'near-line' service with maximum 15 minutes response time.

Keywords: online searching, common command language, viewer, browsing, document delivery, European Patent Office, EPO, full-text searching, dual-mode, near-line, EPOQUE, BNS

1. Introduction

The European Patent Organisation is an inter-governmental organisation set up pursuant to the European Patent Convention which entered into force in 1977. At present, 18 Member States have ratified the Convention (15 of the European Community as well as Liechtenstein, Monaco and Switzerland).

The European Patent Convention establishes an organisation to implement a single procedure for the searching and examination of European patent applications. The executive body of the EPO is the European Patent Office which has its head office in Munich and branch offices in The Hague, Berlin and Vienna. The European Patent Office employs approximately 3700 persons of whom some 2000 are patent examiners (search and substantive examination).

The EPO's main line of business, the delivery of patents, involves the handling and processing of vast amounts of information. Automation is the key answer to dealing with the ever increasing volume of information. The Office is already an extensive user of automation and is expanding its use considerably over the coming years.

Traditionally the EPO patent examiner, for performing a patentability search, has made use of a systematically classified paper collection comprising some 25 MIO patent documents. Over the past years, however, automated tools were installed progressively for supporting electronic documentation and search activities. The present paper gives an overview of the service as now available to close to 2000 end-users, among which the approximately 1200 search examiners are very intensive users.

2. The problem of automating a patent office

One of the important targets in automation projects in the worlds' major patent offices is the field of patent documentation and prior art searching. Indeed, traditional searching methods based on classified paper documents have reached the limits of their capacity. The sheer volume of published documents, currently growing at the rate of around 800,000 new items a year, inevitably leads to longer search times and higher costs, since examiners have more and more information to contend with. Subdividing groups of documents and making

classification more exact can offer only a partial remedy to the problem.

Moreover, more an more we are seeing an ever-increasing number of cross-disciplinary inventions extending over several technical fields at once. Electronic games are one example of this trend. Inventions of this kind mean that more search groups of documents have to be consulted and make classification a more complex task than in the past.

Finally, patent applicants now have access to a wide range of online information systems which are used with great skill. In the majority of cases the ability and the technical know-how to make an accurate assessment of the state of the art before filing the application is at the applicant's disposal. The credibility of patent offices would be called directly into question if their attachment to traditional, i.e. paper methods rendered them unable to obtain results as good as or better than those achieved by applicants themselves.

Patent documents are not just articles containing information: they are highly technical and pose several specific problems: for example they include a large number of drawings, essential to the understanding of their content, and their text relies on a sophisticated and sometimes highly esoteric language and is difficult to read. When designing automated search systems these aspects have to be taken into account. A system based solely on bibliographic, abstract or full-text searching will always be of limited use — text has to be supplemented by graphics. It must be necessary to navigate through the document, moving from a drawing to the relevant part of the text — and vice versa, jumping to another section of the text relating to the same graphical image — without having to scroll through the complete document in order to find the next interesting passage, compare one diagram with another and so forth.

Electronic searching also has to be interactive and follow an iterative rather than a sequential logic. As well as being expensive to set up, automated systems based on sequential consultation of a large number of documents do not and cannot meet the real needs of patent examiners.

Finally, it has to be remembered that searching methods vary in subtle and significant ways from one technical field to another and within a technical field, even from one patent application to another. This is already the case with the traditional paper tools: it is all the more so with their electronic equivalents. The available data, the user-interface and the way to work with the automated tools therefore have to foresee to a large extent the different ways of working that are dependent on the technical field involved (Ref 1).

3. The three basic elements of the EPO solution

The aim of the EPO's automation programme, therefore, was to build an electronic information system whose scope equals or exceeds that of traditional paper documentation. Reviewing the basic requirements of the search examiner and matching them to an electronic search leads to the conclusion that the fundamental process to be supported comprises the following three basic steps (Ref 2):

- identifying the set of documents which are potentially relevant to the patent application at hand (typically 200 documents);
- eliminating irrelevant documents and selecting relevant documents (typically 20 documents);
- in-depth study of selected documents.

These three steps quite often are repeated interactively several times before reaching the final goal, typically consisting of four or five pertinent documents cited in the search report.

The *EPOQUE suite of applications* is the cornerstone of the automated patent search environment at the EPO. In order to support the three steps described above it consists of three major parts:

- an *online search and retrieval tool*, with growing emphasis on the user's added value input;
- an *online viewer part* providing display and browsing of the full documents for one member per patent family — including the first page abstracts and images — in dual mode with ASCII text and facsimile drawings, in the selection/elimination process;
- the electronic equivalence of the entire paper document collection, providing copies (paper and electronic facsimile) of the complete original documents for in-depth study before citation in the search report.

These three parts will be detailed in the following paragraphs.

4. The EPOQUE search and retrieval tool

The EPOQUE host service was installed on the EPO mainframe computer in 1989 (Ref 3) and has been further developed ever since. It enables the search examiners to interrogate a great number of databases from their workstations.

In a common environment, with a sophisticated user interface, interactive interrogation is possible of databases from three different sources:

- internal (databases loaded on the EPO mainframe);
- external (databases loaded on external commercial hosts);
- personal (databases created by the examiners themselves).

The aim is to have these three retrieval parts permanently available for use with one single standardised way

of interrogating them (common command language). The user interface that permits this includes the functions Internal, External, Personal, Preparation and macros, Download, Function Keys, Language and Data Conversion. A simple click with the mouse or combination of keyboard strokes as well as the 'drag & drop' possibility will activate whatever option the user wants to select.

4.1. Main EPOQUE retrieval functionalities

All the internally loaded databases used for searching and retrieval in the EPO have been standardised as much as possible. For instance, in the whole set of databases the patent, priority or application numbers, dates, document types or classification symbols are always written in the same format with the same field name. This allows easier cluster searching and direct cross-file between databases.

The examiner must be able to connect automatically to external databases loaded on commercial hosts, and to interrogate these using the same EPOQUE command language (language conversion) and using standard data format for patent numbers (data conversion), as well as standard field names (tag conversion).

The personal database tool is designed to allow the creation of specific databases on the users' hard disks, enabling the addition of personal codes, remarks or keywords to the documentation in the examiners' own specific search field. This tool also uses a subset of the EPOQUE language and the same database structures.

Important search aids have been created such as the so-called 'standard functions' which are in fact super macros of commonly occurring questions. They can be used without use of the command language. The standard functions are permanently available for selection on the screen, ready for use, wherever else the user may or may not be connected to.

The workstation software includes offline 'preparations', (search strategies) the ability for users to program macros and associate them to function keys, and the ability to share these preparations. The continuous creation of the LOG files on the users' workstation is seen as a major feature in the daily use of EPOQUE. Whatever happens in the dialogue between internal, external or personal sessions, everything is written to a file. This feature not only permits going back and forth in the whole online session while being online but it also allows, in append or overwrite mode, the entirety of sessions executed in the past, at any moment in time, to be consulted.

4.2. Internal databases

The internal databases comprise in the first place those which are produced by the EPO itself. Among these the EPODOC database, which contains the bibliographic data of the EPO's patent documentation as well as abstracts and titles, plays a key role. Many more databases, such as for example EUREG (European Patent Register) and ECLA (European classification scheme), provide valuable information to the search examiner in his or her daily work. EPOS is one of the databases created by the users themselves. EPOS is a collection of synonyms or expressions related to a certain technical concept and organised per technical field.

In the second place come the databases produced by Trilateral partners, such as PAJ (Patents Abstracts of Japan) which gives access to the English language abstracts of patent applications filed with the Japanese Patent Office. From the US Patent and Trademark Office the EPO receives bibliographic data and full texts of US patents, as well as from various other databases such as UCLA (US classification scheme), while the incoming US abstracts are included in EPODOC.

In the area of full-text patent databases the EPO is building up a set of databases covering the so-called PCT-minimum documentation published after 1970 (one patent per patent family). Of this collection more than 2.5 MIO documents are already available in text coded form as they were produced via an automated printing process. The missing 1.3 MIO documents are obtained via an OCR conversion process (see Section 5.2).

Thirdly, the EPO has loaded on its internal host some databases delivered by external producers, currently comprising:

- WPI (World Patent Index produced by DERWENT);
- INSPEC (Information Services for Physics, Electronics and Computing produced by IEE The Institute of Electrical Engineers);
- TDB (Technical Disclosure Bulletin produced by IBM).

More databases in this category will follow in the near future, among which a series of full-text journals published by ELSEVIER is worth mentioning.

4.3. FIRST PAGE — first step to access the full document

In the patent world, almost all patents now have a standardised first page containing bibliographic data, title, abstract and a 'pertinent' image.

The FIRST PAGE software was able to reconstruct on the workstation the first page of the patent for display on the screen, from the abstracts and bibliographic data available in EPODOC and WPI and the 'clipped image' of the front page. FIRST PAGE became available in 1992 but with the recent introduction of the EPOQUE VIEWER system the first page information is available in this new model at the same level as the complete text and all drawings of the full documents, thereby reconstructing the complete application in dual mode, as detailed in the next section.

5. The EPOQUE VIEWER — a browser and navigation tool

During 1995 the EPOQUE VIEWER came into service. It provides display of the full documents including the first page abstracts and images, for one member per patent family in the selection/elimination process. This online dual mode access uses the ASCII form of the full text of the applications (contained in full-text databases) together with facsimile images of the drawing pages (contained in image databases), thereby enabling examiners to eliminate documents from a result list and/or tag documents for further, deeper study either on paper or on screen via the BNS (see Section 6.1).

The success of this patent information system is dependent on whether the viewer part is capable of solving the users' needs when looking at complete documents on screen, which is always a painful task to perform. In this respect every examiner is able to combine the single relevant items of information contained in a full document that he or she wants to see. This combination may always be the same (e.g. drawings and description of the drawings) but it often changes case-by-case.

It is also important to enhance the users' efficiency compared to paper searches by giving added value to full document browsing. Functionalities which cannot be dealt with in conventional paper searches are focusing on different levels of highlighting, real navigation inside a document and, as close as possible, automatic positioning at a needed passage. Allowing the user to mark relevant areas inside the documents is an essential part of this added value navigation (Ref 4).

5.1. EPOQUE VIEWER functionalities

The user may choose which data to display in patent and non-patent literature documents: abstract and bibliographic data, first page image, the full patent document, patent claims and/or drawing pages — all this is made highly flexible. All combinations of all the different pieces of information are possible.

The search for the unique family member present in the EPOQUE VIEWER is for the system to take care off: a list of PNs given by the user is sorted and de-duplicated, and the corresponding documents (only one per family of identical priorities) present in the VIEWER are searched.

The display of the requested data (browsing through full text an drawing images) is done by total interaction of the user with the system. The workstation takes care of the necessary requests for the selected pieces of text and/or drawing image pages which are stored centrally.

Additional string searching is offered through a dialogue box 'HIGHLIGHT ... string of text', repeated automatically when more text comes in. The retrieved strings of text are highlighted and it is possible to jump from one highlight to the next.

The tagging functionality helps out in the way users eliminate or keep documents from a result list. Some examiners go through the whole set quickly just to get a rough idea of the possible pertinence of the documents and then come back to look at a number of pre-selected documents again. Others study every document more thoroughly from the very beginning and therefore need much more detail because the pertinence of the documents looked at may influence the progress of the search.

In order to add value — by the examiners themselves — to the navigation aspects of EPOQUE VIEWER, marking is possible in the full text, especially in the description and in the claims, by adding the marking word immediately before the sentence in the record. This feature is completely controlled by the end-users who are able to see immediately on the screen the markers they have added, modified or deleted.

5.2. Coverage of the EPOQUE VIEWER document collections

The EPOQUE VIEWER collections will cover as much as possible the information needed by the examiners to carry out the selection/elimination process of the search procedure. For the documents published after 1970, EPOQUE VIEWER provides access systematically to the full text and images of the so-called minimum-PCT documents. Per patent family, only one document is available: approximately 4 million unique documents. The same family member provides the text, the drawing images and the FPAGE information (abstracts from EPODOC, WPI, PAJ and clipped image of the first page). Those documents that have text available in ASCII form are loaded immediately. The others are captured by OCR (Optical Character Recognition) techniques of the corresponding BACON facsimile copies. A three-year contract for OCRing the documents needed to fill up the backfile is ongoing.

6. The EPOQUE-BNS viewer — the access to the original document

The third main part of the EPOQUE suite is the BNS application (BACON Numerical Service). This electronic equivalent of the entire numerical collection provides copies (paper and electronic facsimile) of the complete documents for in-depth study before citation in the search report. It is expected that from the combination of an EPOQUE search and VIEWER elimination and selection process, on average only 20 documents will remain in the final 'hitlist'. The BNS provides the examiner with a copy of those 20 documents (in display on screen or in paper form).

6.1. The BNS service

The BNS service which became available early 1996 is not an interactive online service: it works in so-called near-line mode offering an average response time to the end-user of approximately 3-– minutes (the contractually agreed maximum response time is 15 minutes for 98% of the requests). Today the BNS is delivering copies of about 16,000 documents per day. This number includes near-line access (around 3000 documents per day) as well as additional services performed as batch processes, such as the production of extra copies of documents cited in the search reports for mailing to the patent applicants, copies for documentation purposes (8000 documents per day), copies for National Offices of the member states of the EPO, and so on

The data included in the BNS currently consist of almost 25 million documents in numerical order containing about 250 million scanned page-images compressed according to CCITT Group 4 T.6 representing a total volume of around 14 terabytes. The yearly increase is 3–4% of the total volume (approximately 800,000 documents). The data are stored on double length, double density cartridges of 800 MB each, randomly placed in three robots, each containing 6000 cartridges. An active backup duplicates this entire collection in three additional robots (in a separate room for security reasons), with all the tapes organised in reverse order for faster access.

6.2. The BNS usage

As the third tool in the process of online retrieval and display the BNS will deliver the 'original' document to the examiner within maximum 15 minutes, thus allowing the users to keep concentrated on the subject matter which was being analysed during the selection phase.

The use of the BNS workstation interface is straightforward. A working list is created from the tagged PNs in the EPOQUE Viewer and a request sent out to the BNS server. Because of the near-line access mode, several working lists may be issued. Once the BNS server has retrieved the needed documents from the robots, they are copied to a buffer and from there downloaded to the workstations.

The user can consult an 'overview window' and get information about the status of the issued requests. When a request is in the 'ready' status the documents can be displayed in their original form, by browsing from page to page or switching between different parts (description, figures, claims bibliographic information, amendments, etc.) within a patent publication. Zooming facilities and double window display (for comparison between two different pages on one screen) are available.

A major component of the BNS system is the BPS, the BACON Printing Service. Tagged pages, whole documents with tagged pages or specific parts of documents can be printed either on decentralised departmental laser printers or on centralised printers. The priority for the different print batches depends on the number of pages requested to be printed. For detailed in-depth study before the citation in the search report, examiners print approximately 10 documents or parts of documents. The BNS also allows overlays, i.e. electronic labels with information from databases and integration in the printing job of documents from other sources like search reports.

7. Conclusion

The EPOQUE suite of applications represents an important achievement in the area of online patent searching at the EPO, bringing together all major patent and some non-patent literature databases in the world in one standard system. The workstation software is setting new standards of efficiency and comfort for online end-users. With the combination of EPOQUE RETRIEVAL, VIEWER and BNS, the EPO can justifiably claim to have one of the world's most advanced set of tools for automated patent searching.

In order to ensure that the potential of the new tool is fully exploited, the EPO launched two important projects involving all search examiners: the 'learning process' with the goal of refining and perfecting the new working methods, and the 'bottom-up approach' working as a continuous programme to match the different tools and their contents to the needs of the examiners working in different technical fields. The important challenge, indeed, facing the EPO is the management of the evolution of the culture — the creation of new methods of work, their implementation and the adjustments that are necessary at the individual level: the most difficult of all to change but the most vital to maintaining an efficient and high quality patent search.

C. Jonckheere
European Patent Office
Patentlaan 2
PB 5818
2280 HV Rijswijk (ZH)
The Netherlands
Tel: +31 70 3402564
Fax: +31 70 3403320

References

[1] Michel, J. and K. Douglas (1994) The future of search in the European patent system, *World Patent Information*, **16**(4), 193–197.

[2] Van Thielen, J. (1994) Search tools at the European Patent Office: a new approach, *World Patent Information*, **16**(3), 137–139.

[3] Jonckheere, C. (1990) EPOQUE (EPO Query service): the inhouse host computer of the European Patent Office, *World Patent Information*, **12**, 155–157.

[4] Nuyts, A. (1995) A new dimension in patent indexing and searching — the use of navigation tools, *Proceedings of the 1995 International Chemical Information Conference*, Nimes, France, 23–25 October, pp. 78–88.